Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Mass Spectrometry 240 (2005) 195–212
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Ultra-sensitive, high throughput and quantitative proteomics measurements

Jon M. Jacobs, Matthew E. Monroe, Wei-Jun Qian, Yufeng Shen, Gordon A. Anderson, Richard D. Smith*

Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99354, USA

Received 17 August 2004; accepted 24 September 2004
Available online 26 November 2004

Abstract

We describe the broad basis and application of an approach for very high throughput, ultra-sensitive, and quantitative proteomic measurements based upon the use of ultra-high performance separations and mass spectrometry (MS). An overview of the accurate mass and time (AMT) tag approach and a description of the incorporated data analysis pipeline necessary for efficient proteomic studies are presented. Adjunct technologies, including stable-isotope labeling methodologies and improvements in the utilization of liquid chromatography (LC)–MS peak intensity information for quantitative purposes are also discussed. Related areas include the use of automated sample handling for improving analysis reproducibility, methods for using information from the separation for more confident peptide peak identification, and the utilization of smaller diameter capillary columns having lower volumetric flow rates to increase electrospray ionization efficiency and allow for more predictable and quantitative results. The developments are illustrated in the context of studies of complex biological systems.

Keywords: AMT; Proteomic measurements; LC–MS

1. Introduction

With recent advances in genome sequencing, the biological research paradigm is rapidly transitioning towards incorporating an understanding of biology that benefits from a global "systems" perspective. As a result, biology is evolving from a largely qualitative descriptive science to a quantitative and ultimately predictive science in which the ability to collect and productively use large amounts of biological data is crucial. Developing a systems-level understanding of how an organism functions benefits greatly from global measurements of proteins because of their primary role in nearly all cellular processes. The interest in proteome-wide measurements ranges e.g. from the analysis of human plasma for the identification of diagnostic biomarkers, to the characterization of biochemical pathways of microorganisms important for environmental bioremediation or the understanding of human host-microbial pathogen interactions.

Proteomics measurements that yield insight into biochemical processes can lead us to new global predictive computational models that provide a more solid basis for understanding environmental and human health. However, to successfully reach this stage of predictive modeling, major advances in the ability to measure these highly complex systems are still required. It is expected that numerous proteomics measurements (e.g., time course, comparative disease states, a range of environmental perturbations, etc.) will be needed to provide sufficient data for extracting understandings of even the simplest of biological systems that can involve many thousands of different gene products. mRNA expression studies using microarrays have similarly shown that hundreds or thousands of measurements are often essential to support even relatively modest scientific objectives. In addition to throughput, data quality is highly important since the more quantitative and reproducible the measure-

* Corresponding author.

1387-3806/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2004.09.024

ments, the fewer measurements are needed to achieve a given objective.

At present, most global proteomics measurements are qualitative in nature, providing little more than "parts lists" of proteins with uncertain quality and limited information on co- and post-translational modifications. While these more qualitative proteome measurements can be useful, they generally have significant limitations. First, the likelihood that both "false positive" and "false negative" identifications will result from these measurements is substantial; their levels of confidence are often ill defined, and provided only in qualitative terms. These uncertainties are greatest for lower abundance proteins, where measurement quality (e.g., the signal to noise ratio) is lower or where related factors (e.g., the identification of only a single tryptic peptide for a protein) degrade confidence in the identification, but again in a poorly defined manner. Second, because protein detection depends significantly on the sensitivity (and other details) of a specific measurement, relatively small run-to-run variations in detection limits and other aspects of sample handling and instrument performance can result in significant changes in the proteins detected. Important biological processes potentially associated with changes in protein abundances may be obscured by measurement noise, and extensive sets of replicates may be needed to achieve acceptable levels of confidence in such cases.

2. Proteome analysis technologies

Proteome measurement capabilities with the desired comprehensive quality and quantity clearly require further advances in measurement throughput and data quality. The most mature proteome analysis technology is based on separations using two-dimensional polyacrylamide gel electrophoresis (2D PAGE), in conjunction with protein identification using mass spectrometric (MS) analysis and available protein or genome sequence data [1]. However, proteome coverage with 2D PAGE is problematic for proteins that have low or high isoelectric points (p*I* <3.5 and >9.5), are located in membranes, have an extreme range of molecular weights, or that are present in low abundance. The number of 2D PAGE "spots" are often poorly correlated with the number of different proteins identified; those proteins that are identified are predominantly in higher abundance [2]. In addition, the sensitivity of 2D PAGE is generally limited by the need to first visualize the protein on the gel (femtomole levels), [3,4] as well as by the additional efforts required for characterization. Many important regulatory proteins are expressed at such low levels (e.g., <1000 copies per cell) that detection is precluded unless steps, such as extensively fractionating large protein quantities and/or pooling large numbers of gels for processing, are taken prior to 2D PAGE.

Currently, the most effective approaches for protein characterization by mass spectrometry involve digesting the protein to smaller peptide fragments, i.e., a "bottom-up" approach, that's success, is due to the fact that nearly all proteins yield sets of distinctive peptides. One approach often used with 2D PAGE is referred to as peptide mass fingerprinting. A set of peptide fragments unique to each protein is created by digestion and used as a "fingerprint" to identify the protein [5–9]. The peptide mass fingerprinting approach has been broadly applied, but is limited by throughput because single proteins (or a simple mixture) must be isolated prior to MS analyses. Another approach to protein identification is based on information obtained from the dissociation (e.g., using tandem MS) of one or more peptides [10–16]. Tandem MS (or MS/MS) dissociation of only one peptide can be sufficient for identification of the parent protein [17–21], which enables the identification of multiple proteins in a mixture [22].

For protein analysis, the peptide MS/MS approach can be combined with peptide mixture separations prior to MS analysis [23]. The better the overall separation, the more complex the mixture that can be addressed. The separation quality needed to provide a given level of proteome coverage depends on the MS platform applied and the details of the approach.

The first implementation of a combined liquid chromatography (LC)–MS approach for global proteomics was by Yates and co-workers, who utilized two-dimensional (2D) liquid chromatography separations (e.g., MudPIT [22,24]) involving repeated (∼15) sequential step elutions of peptides from a "global" tryptic digest from a strong cation exchange (SCX) resin followed by reversed-phase solvent gradient separations inserted between each salt elution step. The separations were performed in combination with "data-dependent" analysis in which a few of the most abundant peptides detected every few seconds were additionally subjected to MS/MS analysis using an ion trap mass spectrometer. Washburn et al. [22] reported identification of 1484 proteins from *S. cerevisiae* with this approach, which represented 24% coverage of the predicted open reading frames (ORFs). Proteins were identified from all known sub-cellular compartments, with a wide range of functional classifications, and over a dynamic range of ∼10,000 [24]. Although some bias toward detection of very abundant proteins was noted, a large fraction (53%) of the identified yeast proteins had codon adaptation indices below 0.2, indicating that relatively low abundance proteins could also be detected.

While the "shotgun" approach to proteomics developed by Yates and co-workers was a major advancement, the throughput for such MS/MS experiments is still limiting. MS/MS measurements require that peptides be selected one at a time for analysis; thus, while many peptides can be detected in a single first-stage mass spectrum, this approach requires that these peptides be individually selected for a second MS stage analysis. New linear ion trap MS instrumentation has somewhat improved the efficiency of MS/MS analyses, but one analytical cycle is still needed at a minimum for the identification of each peptide. The challenge for complex proteomics samples is that often large numbers of peptides co-elute from even the best LC/LC separations, and the needed MS/MS analysis dictates that only a small subset of these peptides can

be "picked" for the second stage MS fragmentation and identification. Thus, one is faced with a "too many peaks, too little time" under-sampling problem. This problem can be theoretically addressed by even more extensive pre-MS fractionation or separations chromatography to reduce the mixture complexity eluting to the MS at any time point, but this would only come at the expense of significantly lower throughput, increased sample consumption, and likely increased specific losses of peptides.

In general, the "bottom-up" approach to proteomics measurements is highly flexible and comprehensive, and can be adapted and applied to essentially any protein sample. For example, sub-cellular fractions or other sub-proteomes (e.g., the phosphoproteome) can be studied by implementing appropriate sample processing steps [25–31]. The intact-protein "top-down" approaches, where individual proteins are selected for MS analysis without the need for prior chemical or enzymatic proteolysis [32–35], can complement such peptide-level measurements by potentially providing more complete information on modification state(s). However, a significant fraction of proteins remain intractable to intact level analysis due to factors primarily related to their molecular weight and solubility. Thus, while top-down approaches are potentially an important complement to bottom-up approaches, at this time, they are much less mature, and provide much less coverage.

3. Accurate mass and time tag approach

To overcome the "too many peaks, too little time" bottleneck we developed a strategy that increases throughput by avoiding *routine* MS/MS measurements. The technical foundation for this strategy involves advanced separations combined with very accurate mass spectrometric measurements; in particular, ultra-high pressure capillary LC combined with Fourier transform ion cyclotron resonance (FTICR) mass spectrometry, and a supporting data analysis and management infrastructure.

The use of FTICR presently provides a number of advantages over conventional MS platforms, including greater confidence in protein identification and enhanced sensitivity for low abundance proteins [36–39]. Mass spectra can be acquired with resolution in excess of 10^5 and with low to sub part-per-million (ppm) mass measurement accuracy (MMA) [40,41]. These measurement qualities are important because they allow more complex mixtures to be characterized. High quality separations increase the extent of proteome coverage since this depends substantially on the achievable dynamic range of the measurements. With MS platforms such as FTICR that have a maximum ion trap capacity, the overall dynamic range achievable depends significantly on the resolution and peak capacity of the on-line separations used. While the dynamic range achieved in any one spectrum is typically $\sim 10^3$, the effective dynamic range achievable in proteome measurements with combined LC-FTICR improves (has been shown to be $>10^5$) [42] as high abundance species are separated from low abundance species prior to the MS analysis by the ability to regulate ion populations in the FTICR trap [43].

The proteomics approach we have developed is a variation on the "shotgun" proteomics approaches that yields improved sensitivity and throughput. As with shotgun approaches, proteins are cleaved into peptide fragments (e.g., by a specific proteolytic enzyme such as trypsin) after cell lysis and sample processing, producing tens to hundreds of peptides for each protein, and perhaps $>10^5$ in total (depending on proteome complexity and achievable detection limits). These peptides are subsequently analyzed by high resolution capillary LC-FTICR. The capillary LC-FTICR analysis can be preceded by additional sample fractionation, which allows for more complex proteomes to be studied with greater specificity in detection and generally greater dynamic range. However, to avoid the increase in sample analysis and the subsequent lower throughput that would result from additional sample fractionation, when sufficient we use a single high quality separation stage. We do use shotgun LC–MS/MS proteome analyses for the initial peptide identification but importantly, these multiple analyses need only be performed once for a particular biological system which then effectively serve to create a "look-up table" that contains a characteristic accurate mass and LC separation elution time for each peptide, which becomes a unique 2D marker for its identification. These peptide "mass and time" (MT) tags can then be used with subsequent high accuracy mass and normalized LC elution time measurements to identify many peptides in each spectrum without MS/MS, providing both greater sensitivity and increased throughput. The greater sensitivity provides improved identification of lower abundance peptides, and thus better proteome coverage.

The overall proteome measurement strategy has two parts: (1) generating potential MT tags (i.e., peptide markers) from extensive automated LC–MS/MS analyses and (2) high throughput LC-FTICR studies that use the MT tag database in subsequent studies to identify peptides and proteins (see Fig. 1). The peptide MT tags are assigned using conventional software tools (e.g., SEQUEST [44]), providing both the calculated accurate mass (if the identification is correct) and the normalized LC elution time (to $< \sim 2\%$ uncertainty based on present approaches) [44]. Subsequent identification by LC-FTICR accurate mass measurements effectively validates an MT tag as an "accurate mass and time" (AMT) tag. This paradigm applies broadly; e.g., modified peptides or intact proteins can also serve as MT tags. Conceptually the AMT tag approach is similar to the identification of a protein spot in 2D-PAGE studies: once a protein spot has been identified for a particular biological system, a spot at the exact same location in subsequent 2D-PAGE analyses of the same system can generally be assigned with high confidence. The ability to confidently "call" such peptides in an LC-FTICR analysis depends strongly on the specificity provided by the separation combined with the accuracy of the mass measurements.

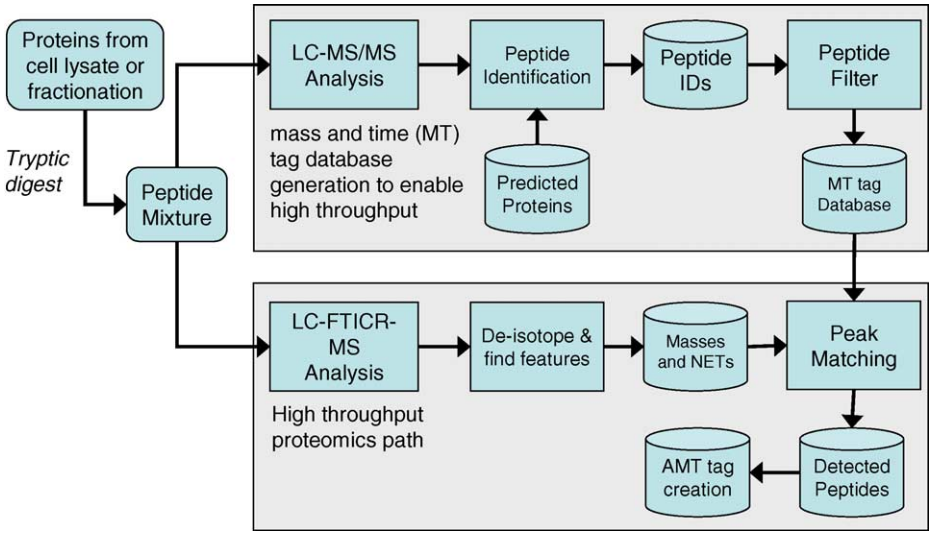


Fig. 1. Basic schematic of the accurate mass and time (AMT) tag approach. The AMT approach consists of two main components. (1) Creation of a mass and time (MT) tag database using a LC–MS/MS peptide identification strategy and (2) use of LC-FTICR–MS for high-throughput accurate mass measurements which will be compared against the MT tag database for the identification of peptides and creation of an accurate mass and time (AMT) tag.

4. Proteomic data analysis pipeline

Proteomic analysis of biological samples—identifying peptides and proteins, and quantifying their abundances—using MS technology generally produces large volumes of data. An analysis often provides thousands of separate MS or MS/MS spectra during the LC separation steps. Data analysis tools are used for performing database searches to identify peptides (e.g. from MS/MS datasets or AMT tags), interpret and extract detected masses from MS datasets, and assign peptide identifications to MS detected masses. These complex multistage analyses also require tracking of experimental conditions and sample pedigree. Additionally, quality control measures are desirable at several stages of the processing to ensure instrument performance and sample preparation quality.

Our laboratory uses an in-house developed proteomics research information system and management (PRISM) to store, track the history of, and provide increasingly automated analyses of proteomic data. PRISM is composed of distributed software components that operate cooperatively on several commercially available computer systems and communicate by means of standard network connections. The system not only collects data files directly from all mass spectrometers in our laboratory, but also manages the storage and tracking of these data files, as well as automating data processing into both intermediate results and final products. Additionally, PRISM collects and maintains meta information about the biological samples used in research experiments and the laboratory protocols and procedures used to prepare them.

PRISM is comprised of two major subsystems: the data management system (DMS) and the mass tag system (MTS) (see Fig. 2). DMS uses various entities to track both data and analysis "results" files, as well as coordinate data handling and maintain contextual background information, e.g., cell culture and stress conditions, sample digestion and processing conditions, experimental factors such as fractionation or chromatographic separations used and MS parameters and conditions, and details for previously processed data such as the peptide identification parameters and criteria used. These entities are arranged in a hierarchical fashion and can be accessed through a web browser interface.

The MTS produces, compiles, and maintains MT tag databases developed in the course of biological studies. Creation of an MT database is typically based upon an extensive series of shotgun LC–MS/MS analyses performed using a variety of sample fractionation, or cell growth and/or treatment conditions. After the peptides are identified from the MS/MS spectra using tools such as SEQUEST the corresponding LC elution time is documented for each peptide. We use an elution normalization algorithm to place all peptides on the same elution time scale, and to correct (normalize) for any small run-to-run variations in absolute peptide elution times. (This step requires that we rigorously use a standard separation to provide high precision elution time information and facilitate a reduction of variation due to run-to-run variability in the LC separation process.) The peptide identifications and normalized elution times are then saved in a relational database along with the appropriate tracking data. The contents of each MT database are determined by selection criteria that are defined in terms of the tracking information that the DMS associates with each analysis job. The associated DMS tracking information is also used inside an MT tag database in order to analyze the effects of e.g. different cell growth conditions on protein expression. MT tag databases are also automatically updated as new analysis results are generated.

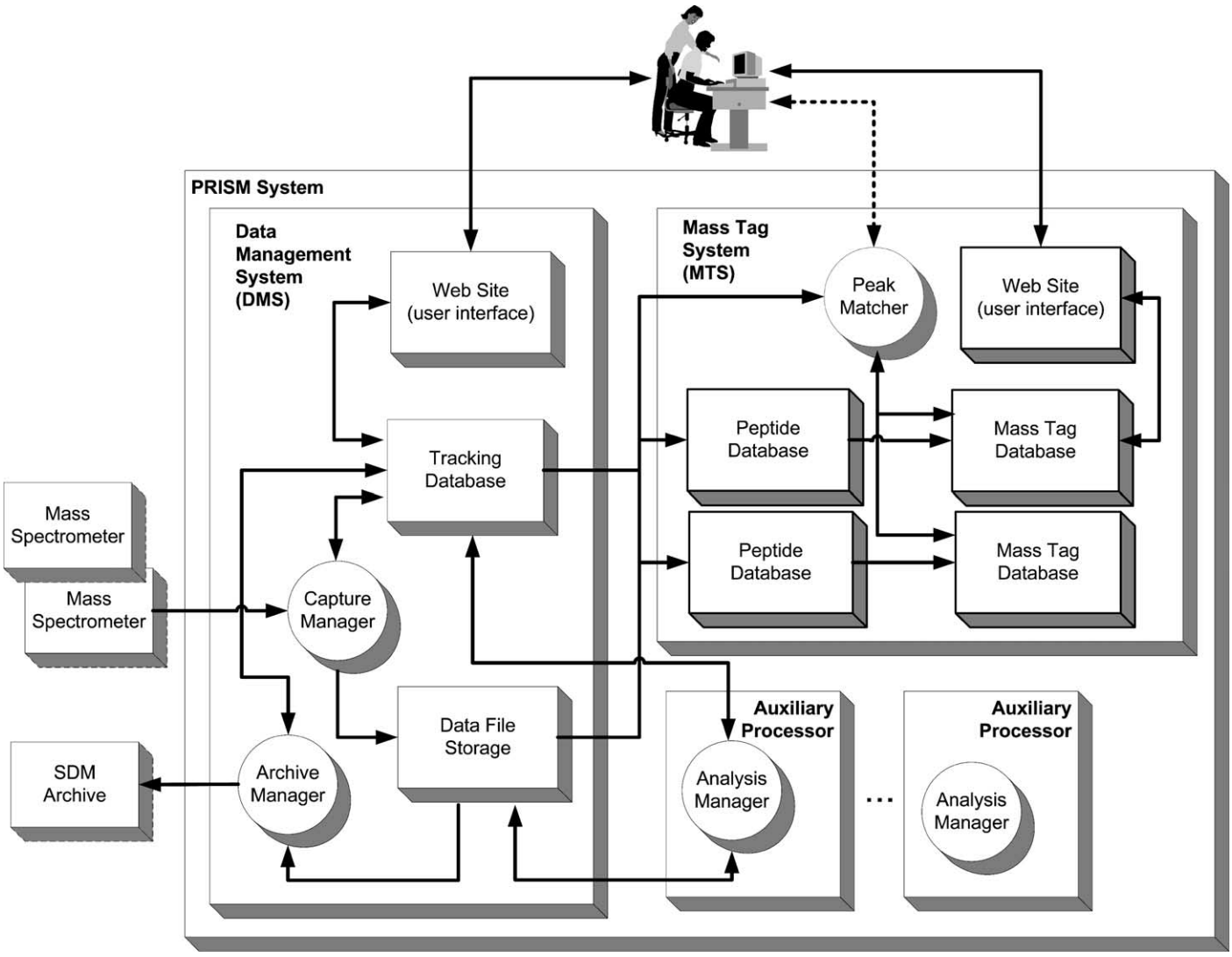


Fig. 2. Basic structure of the proteomics research information system and management (PRISM).

An important step in our proteomics data analysis pipeline involves reducing the large volumes of raw data to a form amenable to all downstream data analysis algorithms. Raw data include spectra from both LC–MS/MS, which are used to search databases of protein sequences and to generate tables of predicted peptide identifications, and LC-FTICR, which are analyzed and converted to tables of separation times and accurate masses that represent individual species detected in each spectrum. The data reduction step for LC-FTICR analyses is significant; for example, a 10 Gb raw data file single experiment using a 9.4 T LC-FTICR is reduced to a table of detected masses that is on the order of 10 Mb in size, a three order of magnitude reduction in volume. The data reduction step for LC–MS/MS datasets is less significant; a typical capillary LC–MS/MS dataset file size from an ion traps is ∼20 Mb and generally yields a list of predicted peptides having a file size of 1 Mb or smaller.

Data reduction for high performance FTICR mass spectrometers, and increasingly for TOF instrumentation, involves processing each spectrum to identify peaks contributing to the isotopic distributions for different species and then to determine the corresponding neutral masses for the detected species. Fig. 3 shows a typical isotopic distribution for a peptide detected with an 11.4 T FTICR; a single spectrum

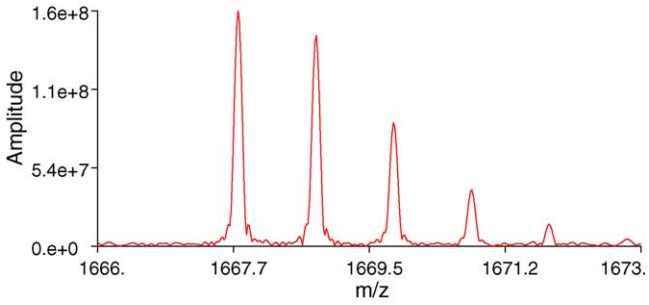


Fig. 3. Typical isotopic distribution for a peptide detected by FTICR. For data reduction, thousands of such isotopic distributions in a single analysis are mass transformed into a table of mono-isotopic masses.

can contain thousands of isotopic distributions similar to the distributions shown in this figure. A typical spectrum is described by using a data vector containing >200,000 points and calibration information specific to the instrument. After processing the spectrum to extract peak information, a table of detected masses and their corresponding intensities is developed for use in later stages of the proteomics data analysis pipeline. The signal is processed to reduce the isotopic distribution for a species described by ∼400 raw data points to a monoisotopic mass e.g. of 1667.9055 Da. The processes of converting the isotopic distributions to tables of masses are referred to as mass transformation and de-isotoping. This capability has been developed in our laboratory and implemented in software named ICR-2LS that utilizes an approach based on the THRASH algorithm [45].

Interfacing between the MT tag database and LC-FTICR datasets is accomplished by the software program VIPER, whose main objective is to connect LC-FTICR datasets with previously established MT tags, using a very restrictive criterion to eliminate or minimize ambiguities. The software performs MS data inspection, visualization, and analysis applications, allowing for rapid display and analysis of large datasets produced by both FTICR and TOF mass spectrometers. VIPER receives a "PEK" file (text file created by the ICR-2LS program) that contains a list of molecular masses and their corresponding intensities extracted from raw spectra and can display the result as a 2D plot with elution time (spectrum number) and molecular mass as coordinate axes. The abundance of the isotopic distribution can be indicated by a variable spot size, and color can be used to code abundance ratios and/or charge states of individual distributions.

In the context of VIPER, a chromatographic peak is considered as a set of measurements for a given species (typically in sequential spectra) and designated as a unique mass class (UMC). In other words, the UMC is an equivalency class of isotopic distributions close in molecular mass and elution time. The VIPER program contains several different highly customizable functions whose purpose is to "populate" a 2D LC–MS peptide landscape with UMCs. The 2D display allows one to look for groups of related data using data clustering algorithms that detect species having similar monoisotopic masses, elution times, intensities, etc. Each group or UMC has a median mass, central LC normalized elution time (NET), and abundance estimate, computed by summing the intensities of the MS peaks that comprise the UMC (see Fig. 4). Since even the highest feasible mass measurement accuracy may be insufficient to confidently choose among possible peptides having similar or identical masses (as often found in a proteome of a complex organism), we also make use of the NET value to reduce ambiguity in the identification phase. Once identified, a UMC can be used to search against MT tag databases for matching peptides, to compare features between related datasets, or to compare features within a dataset. This functionality, together with UMC visualization and reporting capabilities, makes VIPER an invaluable component in the proteomic data pipeline.

A statistically based algorithm using a least-squares method maps the NET and mass of each UMC in an LC-FTICR dataset to the NETs and masses in the MT tag databases. Fig. 5 illustrates a situation in which several MT tags within tolerance of a given UMC (most likely a peptide) result in an ambiguously identified UMC. In this figure, the mass accuracy defines the height and the elution time reproducibility defines the width of the ellipses, which are the acceptable error boundaries of the MT tags. Instrumental or procedural enhancements that improve either the mass accuracy or the NET precision will decrease the number of features with ambiguously matching MT tags. For situations in which several MT tags do match a feature, one can compute a probability of the most likely match based on the standardized squared distance between a given feature's mass and elution time and each MT tag's mass and elution time,[46] as illustrated by the Probability column in Fig. 5. For the example shown, the fourth peptide is the most likely match, because both its probability and its SEQUEST XCorr value are higher than those for the other peptides.

Our technology platform and analysis pipeline for high throughput global proteomic measurements has proven to be

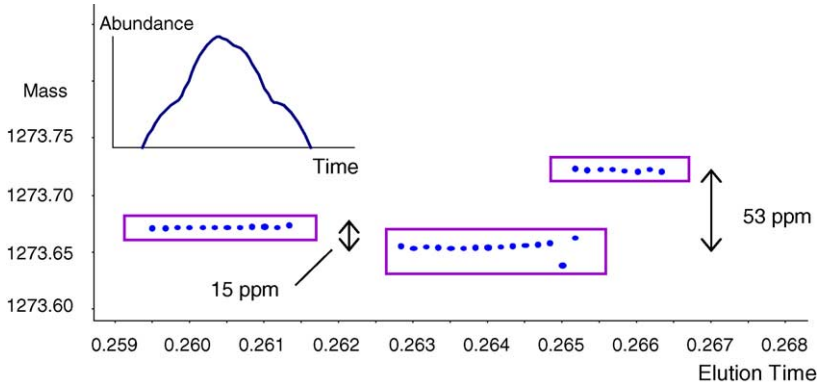
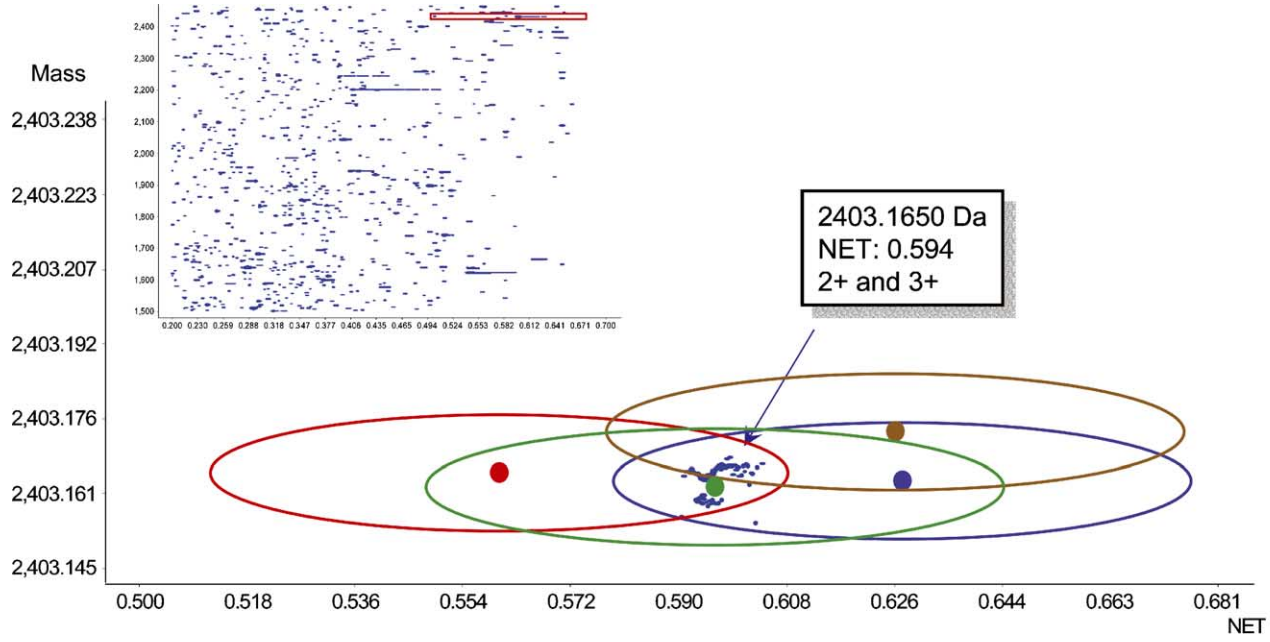


Fig. 4. Visual example of unique mass classes (UMCs) which are 2D representations of a mass tag (peptide) chromatographic peak, see inset. Each spot within the group represents an isotopic distribution, near identical to the others in the group but separated only by slight changes in the elution time representing the elution of a single peak.



Peptide	Monoisotopic Mass	Avg NET	High XCorr	Probability of Match
E.QELRETEHLLHDENEDLRK.L	2403.1777	0.627	2.50	0.05
L.STPTHAILSAANPTSAAGNPGGESVP.-	2403.1665	0.561	2.71	0.22
R.AQEHSLPNFAGGQHFFEYLLV.V	2403.1647	0.631	2.60	0.19
R.MPCAEDYLSVVLNQLCVLHEK.T	2403.1635	0.594	6.17	0.54

Fig. 5. The inset shows a two-dimensional plot with normalized elution time (NET) on the horizontal axis and mass on the vertical axis. The region highlighted in red on the inset plot is illustrated by the larger plot. A single detected feature is present at 2403.1650 Da and 0.594 NET. Four MT tags match this feature with the tolerances dictated by the mass accuracy and elution time accuracy of the instrument. The probability of each match was computed using the standardized squared distance between the feature's mass and elution time and each MT tag's mass and elution time.

broadly effective for identifying proteins from both microbial and mammalian systems. The continued development of this strategy involves targeted analysis to provide a larger increase in throughput and data production that is both quantitative and of improved quality. Also critical is the continued progression of supporting capabilities to make this increased data stream more useful for biological applications, and its extension to both modified peptides and "top–down" approaches.

5. Quantitation strategies

Proteome measurements often involve comparing protein abundances between two cellular populations that differ as a result of some change or perturbation. For comparative studies that employ stable-isotope labeling, the AMT tag strategy can increase throughput and precision by directly comparing two proteomes in the same analysis, such as comparing perturbed systems to a common "reference proteome". A stable-isotope labeled (e.g., ^{15}N or ^{18}O labeled) reference proteome provides an effective internal standard for many peptides [47–49]. A key advantage of this well-established approach is that variations due to sample processing (after mixing) and analysis are eliminated, allowing the relative abundances to be determined to better than 10–20% in many cases [47,50,51].

One stable isotopic labeling method that we have found especially useful for comparing relative protein abundances is via enzyme transfer of ^{18}O from water to the C-terminus of peptides [52]. In this approach, proteins isolated from two samples are separately digested in either ^{16}O or ^{18}O water using trypsin. The oxygen atom, either ^{16}O or ^{18}O, from water is incorporated into the newly formed C-terminus in each tryptic peptide, thus providing an isotope tag for relative quantitation. Initial work with this approach indicated that labeling efficiency can vary somewhat for different peptides, leading some peptides to incorporate two ^{18}O atoms per peptide while others only incorporate one ^{18}O [53,54]. This variability and its repercussions for quantitative analyses made its use problematic. Recently, however, the mechanism for ^{18}O transfer was demonstrated to be an enzyme-catalyzed oxygen exchange reaction [55], allowing much more consistent labeling of two ^{18}O atoms to be obtained during a post-digestion trypsin-catalyzed ^{16}O/^{18}O exchange reaction [56]. The ad-

vantages of the enzymatic ^{18}O labeling approach include the ability to label all types of samples (including tissues, cells, or biological fluids), the simplicity of the reaction and its specificity for C-terminus of tryptic peptides, and identical elution times for both light and heavy isotopic-coded peptides in a pair. The incorporation of two ^{18}O atoms gives a mass difference of 4 Da between the ^{16}O and ^{18}O labeled tryptic peptides, and is most effective when using a high resolution mass analyzer such as TOF or FTICR to effectively resolve the ^{16}O and ^{18}O labeled peptide pair peaks and allow quantitative determination of relative abundances.

Recently, our laboratory has applied the trypsin-catalyzed $^{16}O/^{18}O$ labeling and the AMT tag approach to quantitative studies of relative changes in protein concentrations in treated (lipopolysaccaride) versus untreated human blood plasma samples (Qian et al., in preparation). Initial work involving human plasma samples centered on developing strategies for effective MS analysis and on creating a comprehensive mass tag database for subsequent application of high mass accuracy and quantitative studies [57,58]. More than 600 LC–MS/MS analyses were performed to comprehensively identify peptides for the MT tag database. Multiple approaches were used, including extensive SCX fractionation and depletion of high abundant proteins prior to LC–MS/MS analysis to obtain the necessary dynamic range in protein detection. Fig. 6 shows the chromatographic scheme of a subset of the results (Shen et al., submitted for publication). LC-FTICR analysis was then performed with $^{16}O/^{18}O$ labeled peptide mixtures from plasma. Fig. 7 shows a partial 2D-display of the results from a single LC-FTICR analysis and an example peptide pair, illustrating an increase in the protein's plasma concentration in conjunction with lipopolysaccaride treatment.

Also currently being applied in our laboratory is a Quantitative Cysteine-peptide Enrichment Technology (QCET) [53] that allows high throughput systematic identification and quantitation of proteins expressed in mammalian cells. This approach is an alternative to the isotope-coded affinity tag (ICAT) approach and can be readily applied to proteome-wide measurements of very small samples. The QCET approach combines quantitative $^{16}O/^{18}O$ labeling with specific capture and isolation of cysteine containing peptides, as shown in Fig. 8A. Additional benefits include the reduction in sample complexity that results by the elimination of typically 80–90% of non-Cys containing peptides and the

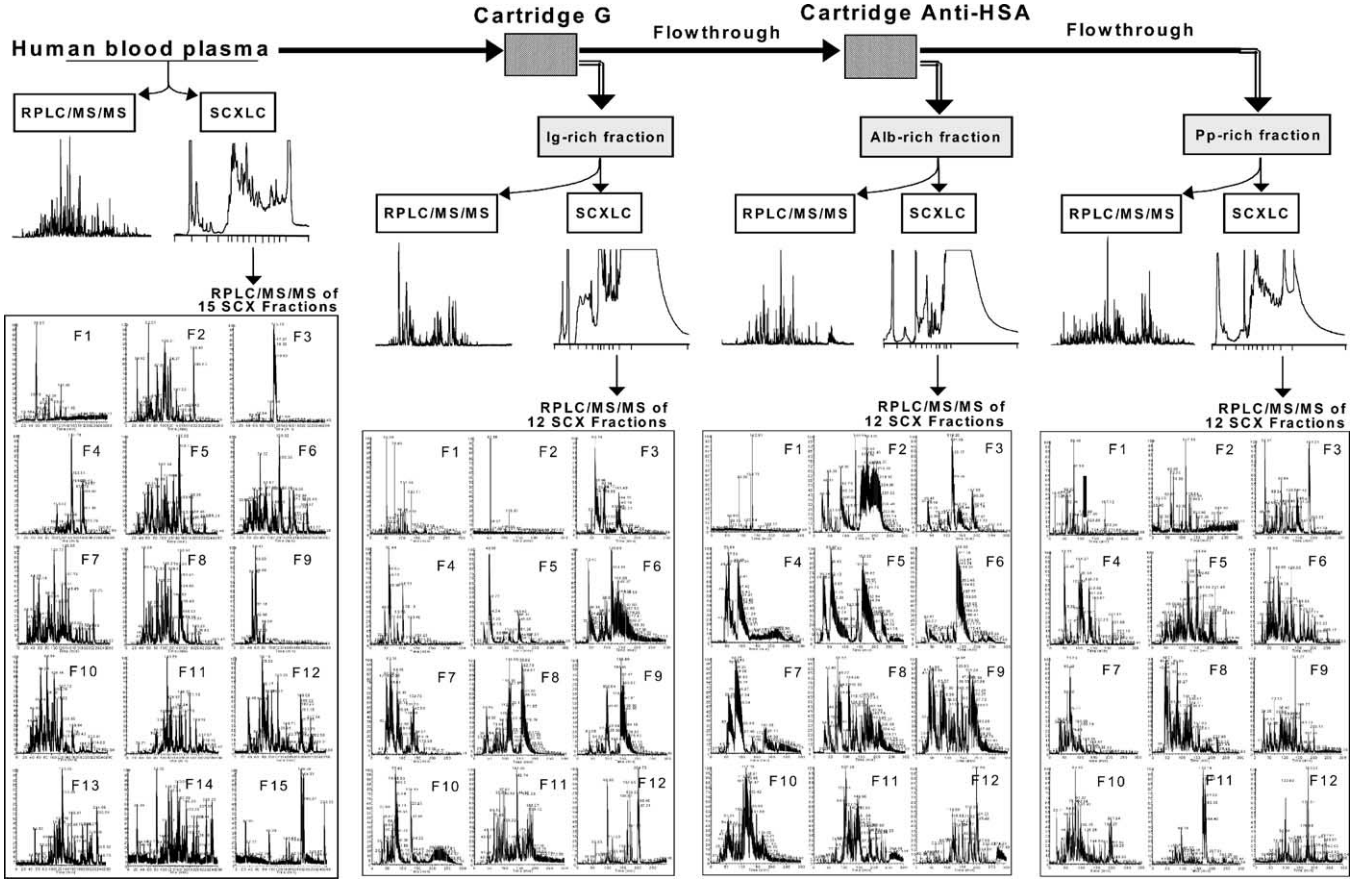


Fig. 6. Chromatographic representation of a pre-analysis separation and depletion scheme of a human plasma sample for RPLC–MS/MS analysis and creation of a mass tag database for further high mass accuracy studies. Depletion was performed at the protein level with the removal of both immunoglobulins and human serum albumin, cartridge G and cartridge anti-HSA, with the various protein fractions submitted to tryptic digestion and strong cation exchange (SCX) separation prior to MS/MS analysis (Shen et al., submitted for publication).

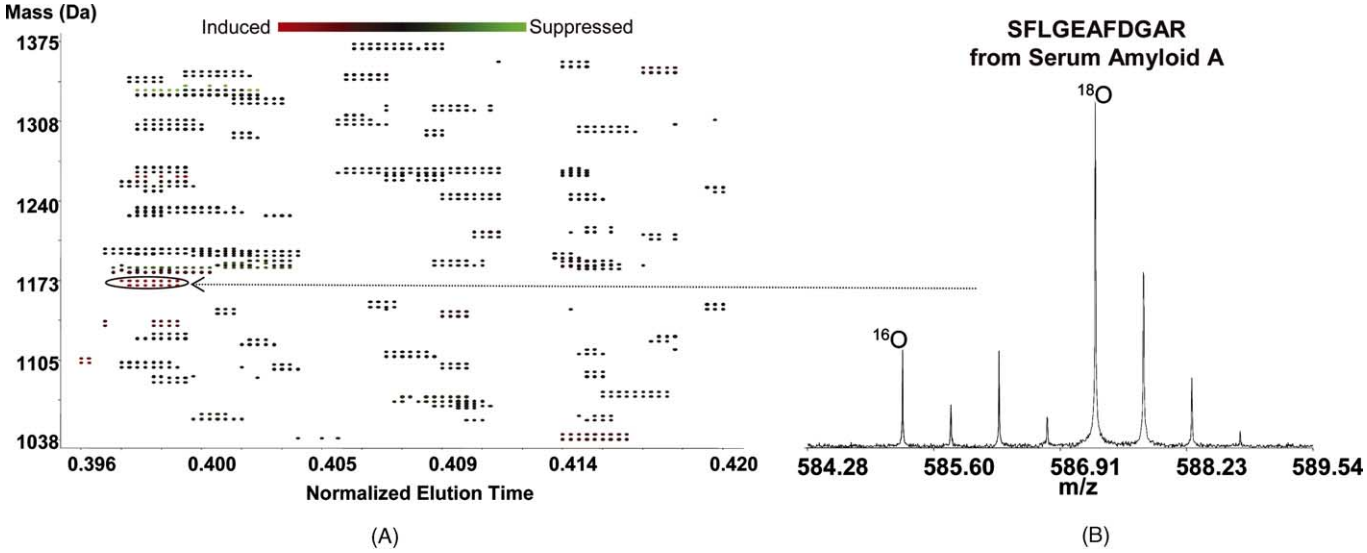


Fig. 7. (A) A 2D plot generated by the software program VIPER showing $^{16}O/^{18}O$ peak intensity pairs using LC-FTICR technology. (B) Example peptide identification of an up-regulated protein by comparing normal plasma and plasma from an LPS-administered subject identified by the AMT tag approach and ^{18}O labeling [86].

subsequent improved peptide identification (essentially all recovered peptides should contain a Cys residue). Using QCET, proteins from two cell states or conditions are prepared and separately digested by trypsin under identical conditions, with tryptic peptides from both samples exclusively labeled with either ^{16}O or ^{18}O by immobilized-trypsin. The differentially labeled peptide samples are then combined, and Cys-peptides are selectively captured using thiol-specific affinity resin (Fig. 8B) and released from the resin by incubating with a low molecular weight thiol. The enriched Cys-peptides are analyzed by LC–MS/MS and LC-FTICR, and identified and quantified using the AMT tag approach.

The quantitative cysteine-peptide enrichment approach was initially applied to study the differential protein expression in human mammary epithelial cells (HMEC) following phorbol 12-myristate 13-acetate (PMA) treatment. To generate the HMEC mass tag database, Cys-peptides enriched from the tryptic digest of cellular lysates were fractionated by SCX chromatography and analyzed by LC–MS/MS. An MT database was generated, containing 6222 identified peptides (covering 3161 proteins). To compare the relative protein abundances between the naïve and PMA-treated cells, 100 μg total protein from each sample was digested, labeled, and combined for Cys-peptide enrichment. The re-

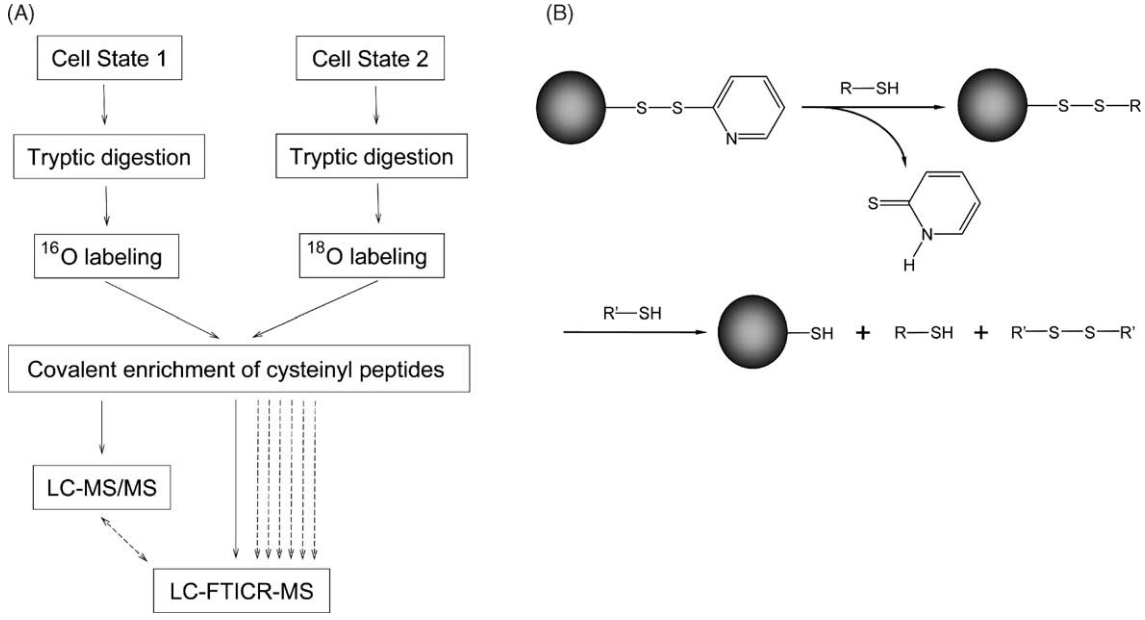


Fig. 8. (A) Strategy for quantitation of differential protein expression. (B) Reaction scheme for covalent capture of a cysteinyl peptide (R-SH) on thiopropyl sepharose 6B. R′-SH represents a low molecular weight thiol such as DTT [53].

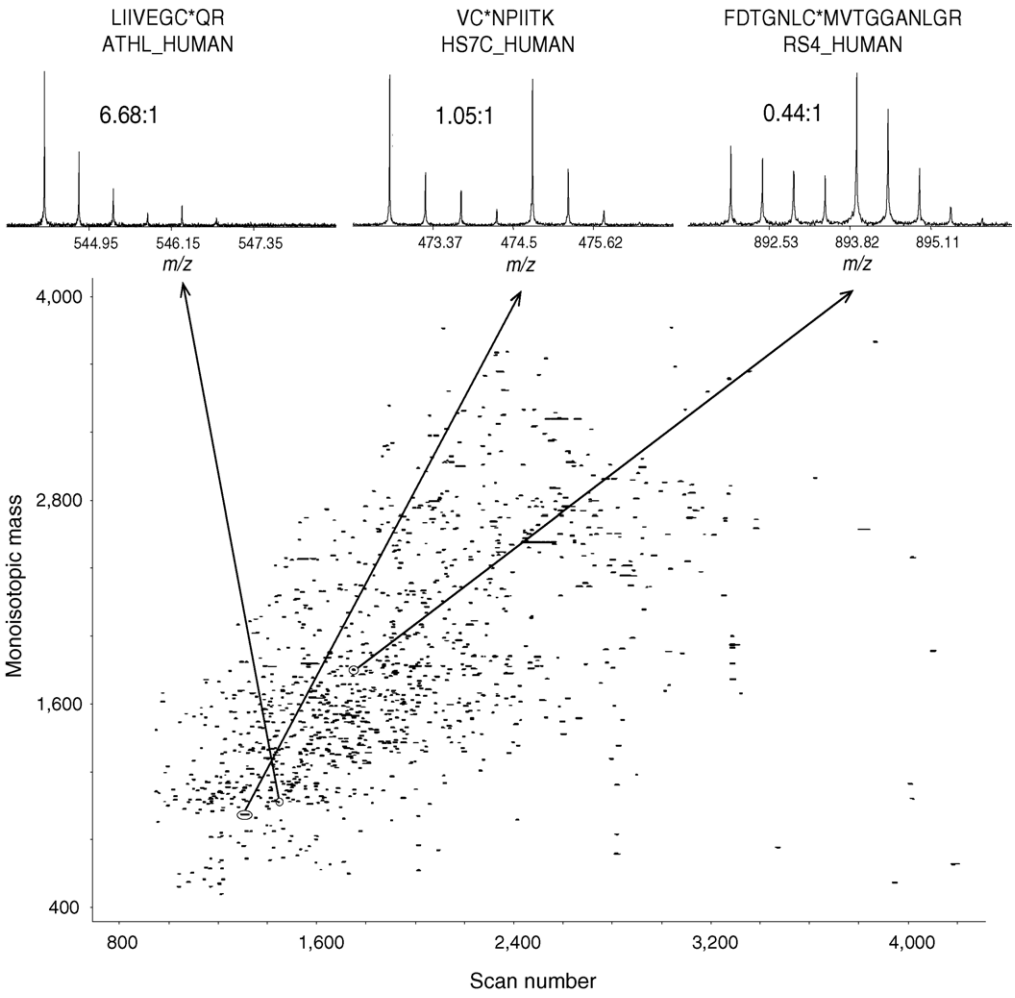


Fig. 9. A 2D display of 1348 peptide pairs from which 935 pairs were identified as unique AMT tags corresponding to 603 proteins identified and quantified. Insets show three examples of peptide pairs with their sequences, corresponding proteins, and the $^{16}O/^{18}O$ ratios [53].

sulting peptide sample was analyzed by LC-FTICR, and a total of 1348 labeled peptide pairs were observed in a single analysis (see Fig. 9). Among these, 935 pairs identified as AMT tags corresponded to 603 proteins [53]. The QCET approach significantly reduces the overall sample complexity by use of high-efficiency Cys-peptide enrichment and makes the AMT tag approach effective for applications to complex mammalian proteome samples. Similarly, the primary disadvantage of the approach is that the number of peptides detected per protein is generally small, but confidence in the protein identifications should remain high due to the added constraint of the identification of only cysteine containing peptides.

6. MS peak intensity based quantitation

Comparative measurements based on isotopic labeling generally require that both versions of the peptide (labeled and unlabeled) be detected, but often large changes in relative protein abundances between two labeled samples result in detection of only one of the peptides (e.g., when there are large abundance changes or low signal to noise levels for the measurements). Approaches based upon the use of peak intensities are attractive for this purpose, but peptide abundance measurements obtained using MS signal intensities may vary significantly for reasons that can include variations in ionization efficiencies and losses during sample preparation and separations. While more readily useful for large differences in abundances between samples, peak intensity measurements have been less effective for studying more subtle variations. However, our studies indicate that proper control of the sample process and analysis conditions (e.g., for electrospray ionization) yields data with high reproducibility between runs and provides a basis for more effective quantitation. Other recent reports have also described the use of peptide peak intensity information for determining changes in relative protein abundances based upon different normalization techniques [59,60]. While less precise than stable-isotope labeling methods, such approaches have the advantage of not requiring additional studies to prepare samples with isotopic labels (or additional sample processing

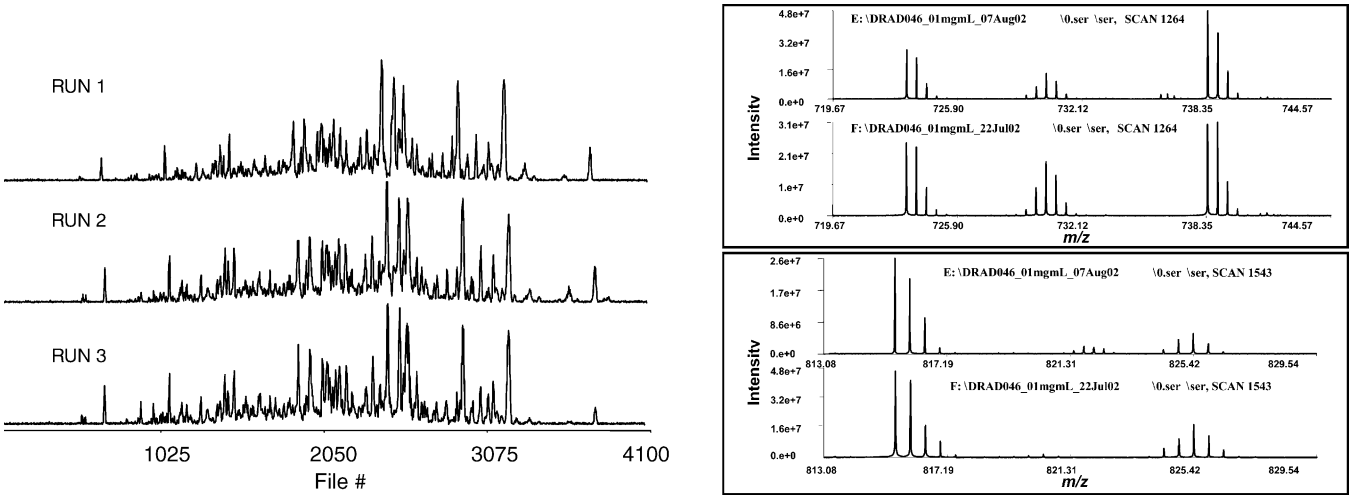


Fig. 10. Left, total signal capillary LC-FTICR chromatograms obtained for three unattended overnight 'back-to-back" analyses of the same tryptic digest of a *D. radiodurans* proteome sample. Right, portions of the mass spectra acquired during two capillary LC-9.4 T FTICR runs conducted 2 weeks apart using automated injections of the same *D. radiodurans* proteome sample [75].

steps to add labels), and are broadly applicable regardless of sample type.

In spite of providing the highest precision likely achievable for comparative analyses, stable isotope labeling approaches have weaknesses that arise when both labeled versions are not detected. Thus, approaches based on MS intensity data can be used to complement isotopic labeling approaches. Issues that dictate the utility of utilizing both approaches are primarily related to: (1) the run to run reproducibility of proteome analyses, (2) the effectiveness of data normalization approaches (similar to those used for microarray data analysis), (3) the linearity of signal response as a function of protein concentration, and (4) other factors that can cause variation in response.

We found that the run-to-run reproducibility of proteome analyses improved dramatically when a fully automated capillary LC-FTICR system [61] was implemented. As an example, Fig. 10 (left) shows the run-to-run reproducibility observed for the total ion chromatograms reconstructed for triplicate LC-FTICR analyses of a *D. radiodurans* proteome sample, and (right) the reproducibility obtained in the mass spectra for an identical sample analysis after 2 weeks. Similarly, Fig. 11 shows the variation in intensities for peptides obtained from six replicate analyses of the same *S. oneidensis* proteome sample. The average coefficient of variance

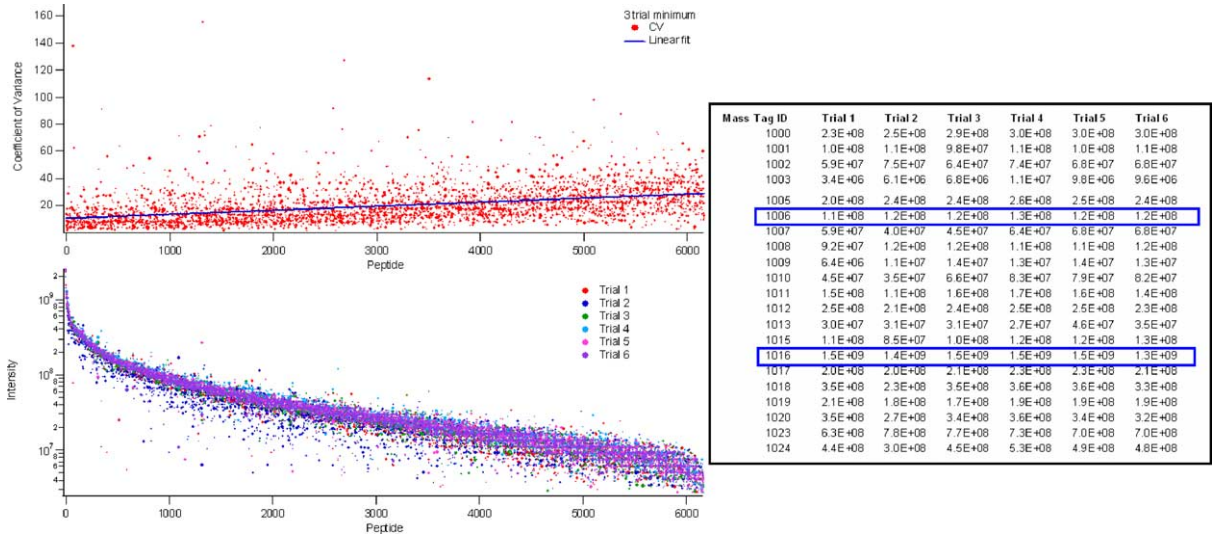


Fig. 11. Left, reproducibility of absolute abundance values for mass and time tags identified in six replicate capillary LC-FTICR analyses of a *S. oneidensis* typically digested proteome sample. The upper plot represents the coefficient of variance in intensity across the six replicates while the lower plot compares the intensity values seen for each identified AMT tag across the replicates. Right, an excerpt from the raw intensity values is shown on the left. The two boxes illustrate the reproducibility of intensity values for a given AMT tag across the 6 replicates without normalization between the replicates.

is ∼10% for the highest abundance peptides and increases to ∼40% for low abundance peptides, illustrating the excellent reproducibility obtainable using automated capillary LC-FTICR.

These results support the feasibility of using quantitation approaches based on intensity data, and are consistent with several other reports [59,60]. However, such approaches have been criticized because of the possibility of ESI "suppression" effects that can arise due to the presence of solution matrix components or peptides eluting at the same point in the separation. Such suppression effects could make determining changes in abundance as well as determining the relative abundances of *different* peptides problematic. Since the extent of suppression is expected to be highly dependent upon the precise solution composition, even comparative measurements for the *same* peptide could be problematic. To address this issue, analyses should be conducted under conditions where ionization suppression effects are minimized (and ideally eliminated). For this reason we now discuss relevant issues related to ESI performance.

7. Electrospray ionization efficiency

The conditions under which ionization suppression occur are relatively well understood and are related to both analyte concentration and ESI volumetric flow rates [62–64]. Large, conventional flow rates result in greater compound-to-compound variation due to the increased analyte competition for charge and the proximity to the electrospray droplet surface. Converting the electrospray ionization from a conventional flow rate to a nanoflow regime (see Fig. 12 [62]) produces smaller initial droplets that allow ions to form more rapidly and efficiently with less heating. Smaller droplets move towards the periphery of the electrospray plume due to their smaller inertia and higher charge density [65] (Fig. 12, top); consequently, conventional electrosprays are generally sampled "off-axis" of the MS inlet, where a majority of the ions are desolvated and the concentration of large problematic droplets or clusters is minimal. Depending on the flow rate, sample concentration, and surface activity, a large portion of the analyte (peptide sample) can form these charged clusters or residue particles at conventional flow rates (μL/min). Also problematic is the limited ∼1 cm spacing between the electrospray emitter and the MS inlet combined with the expansion of the electrospray plume, resulting in a reduction in MS inlet efficiency (Fig. 12, top), as well as limiting sensitivity.

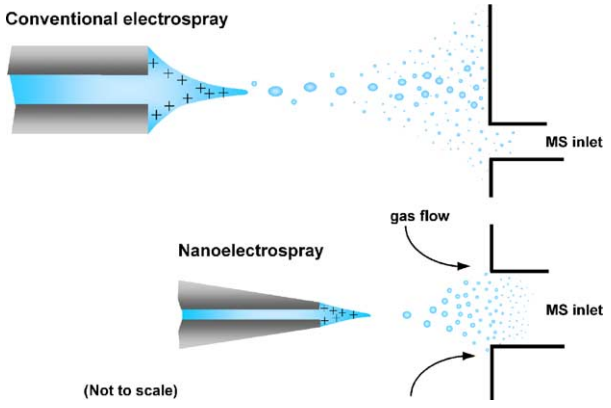


Fig. 12. Normal flow rate electrospray (top) vs. a nano flow rate electrospray (bottom) that produces smaller droplets. By allowing closer proximity to the MS inlet, the lower flow rate electrospray affords more efficient ion introduction (see Table 1) [62].

Observations indicate nano-ESI detection response is more concentration sensitive and less mass sensitive than with conventional flow rates. It is well known that increasing the flow rate in the conventional regime generally does not significantly increase the signal [66]. In our initial work with ESI interfaced to capillary electrophoresis (CE)–MS, we were able to demonstrate sub-femtomole detection limits [67] due to the relatively low CE flow rates. Higher sensitivity was obtained in subsequent studies by employing even smaller diameter capillaries, which serves to provide a reduction in both the sample size used in the analysis as well as the liquid flow rate in the capillary [68]. The use of 10 μm i.d. capillaries for CE–MS [69] provided low nL/min flow rates, and resulted in a mass sensitive response due to the limited delivery of charge carrying species. Later, we described an ESI source that used an etched conical ESI emitter and sub-μL/min flow rates that provided improved stability and utility. The smaller electrospray droplet size enables the emitter to be aligned "on-axis" and closer to the MS inlet, yielding a more efficient ion delivery and increase in sensitivity (Fig. 12, bottom) [70]. Table 1 compares the characteristics of conventional (5 μL/min) flow rates with nano 20 nL/min ESI flow rates. The work performed by both Wilm and Mann [71] and Fernandez de la Mora and Loscertales [72] with respect to characterization of ESI models has provided helpful insight into these relationships.

With sufficiently low flow rates and concentrations delivered to the electrospray there is on average less than one analyte molecule per droplet. Thus, the analyte is dispersed into very small, easily desolvated charged droplets with the ionization efficiency approaching 100% [62]. Even at reduced rates in the intermediate flow levels, ESI efficiencies becomes more uniform (see Figs. 13 and 14) [62], and the efficiency of ion introduction through the MS inlet is enhanced. Another advantage of this mode of operation is that at low flow rates ESI can be used to effectively study compound classes that can be otherwise problematic due to their low surface activity (e.g., highly hydrophilic oligosaccharides) [73,74]. Additionally, the concentration where linear response ends can occur at a fairly sharp boundary [74]. Fig. 15 shows LC-FTICR peak intensities versus sample size for three abundant peptides obtained from LC separation using a 30 μm-i.d. capillary [75]. For sample sizes below a given level, MS peak intensities (or peak areas) are generally in the regime where signals increase linearly with sample size and provide the best quantitation. Since smaller i.d. capillaries provide a linear response over a wider range of sample concentrations, our

Table 1
Comparison of electrospray characteristics at conventional and nano-ESI flow rates [62]

Flow rate	Conventional (5 μL/min)	NanoESI (20 nL/min)
Droplet diameter[a]	1.4 μm[b]–6 μm[c]	150 nm[c]–220 nm[b]
Electrospray current[d]	200 nA	12 nA
Droplet generation rate	7×10^{5}[c]–6×10^{7}[b](1/s)	6×10^{7}[b]–2×10^{8}[c] (1/s)
Molecules/droplet (1 μM analyte concentration)	860[b]–7200[c]	1[c]–3[b]
Charges/droplet	2×10^{4}[b]–2×10^{6}[c]	400[c]–1250[b]
Charges/analyte (1 μM concentration)	25[b,c]	360[b,c]

[a] [74].
[b] [72].
[c] [71].
[d] Measured.

laboratory has implemented a "periodic upgrade" to its high throughput capabilities for implementing decreasing capillary i.d. (and ESI flow rate) from 150 μm to 75 or 50 μm i.d. used at present, after demonstrating robust operation of the smaller capillary i.d. columns using an automated ultra-high pressure LC-ion trap MS system. The ideal approach is to use LC capillary columns that provide the lowest flow rates, while remaining robust enough for high throughput operations. Available results suggest that column i.d.'s extending to as small as 10 μm should be beneficial.

There are challenges associated with obtaining high quality LC separations at nano-flow rates, beginning with the introduction of the sample to the column. Even when using the highest reasonable LC pressures possible, injection times for loading a 10 μL sample onto a 90 cm × 15 μm i.d. column can take upward of 500 min due to the inherent flow resistance of the longer packed capillaries. We have overcome this time-limiting step by developing a system that starts off with a short pre-column that has a much larger i.d. onto which the sample can be quickly loaded. We found that a 4 cm long,

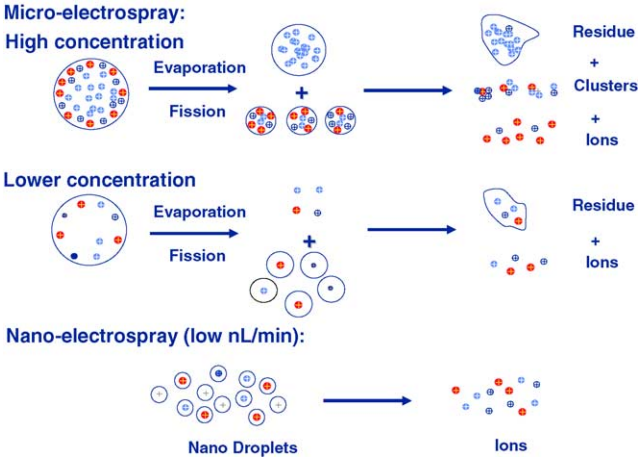


Fig. 13. A simplified illustration showing how concentration and flow rate can affect the ESI process. For larger flow rates, which produce larger droplets (Table 1), analyte surface activity, concentration, and competition from other species can affect overall ionization efficiency, the extent of ionization "suppression" and quantitation. At sufficiently low flow rates and analyte concentrations, each droplet contains on average less than one analyte molecule, ionization efficiency is 100%, and suppression/matrix effects are eliminated [62].

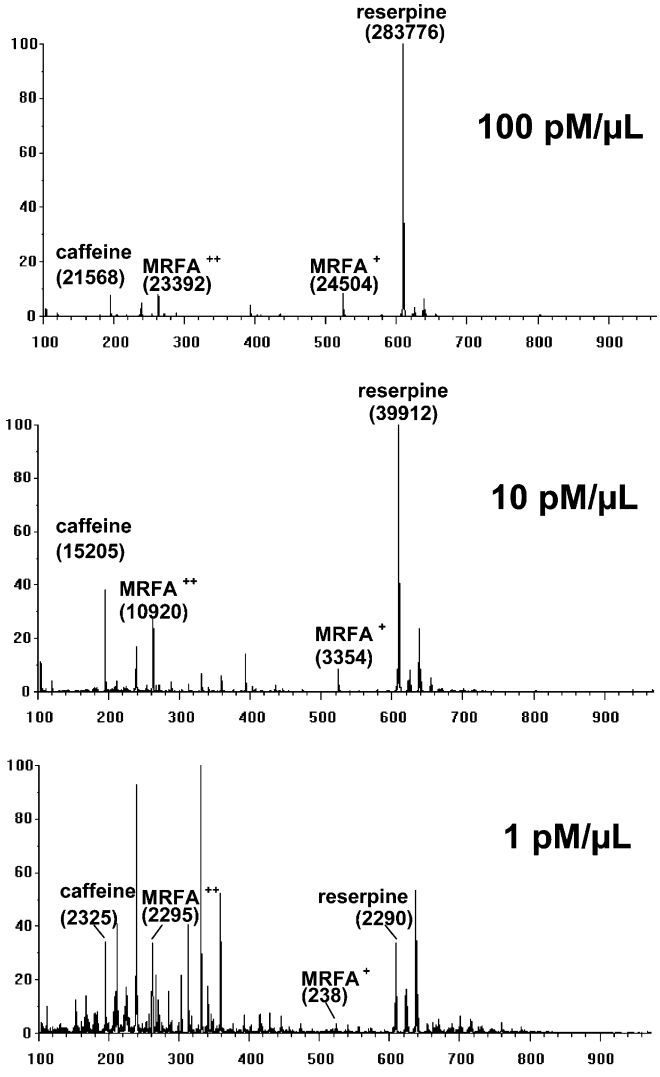


Fig. 14. Mass spectra for a three-component mixture at three concentrations. At the highest concentration (top) reserpine displays greater intensity than other components, but peak intensities become nearly equivalent at the lowest concentration (bottom). As electrospray flow rate is decreased, ionization competition/suppression is avoided at increasingly higher concentrations [62].

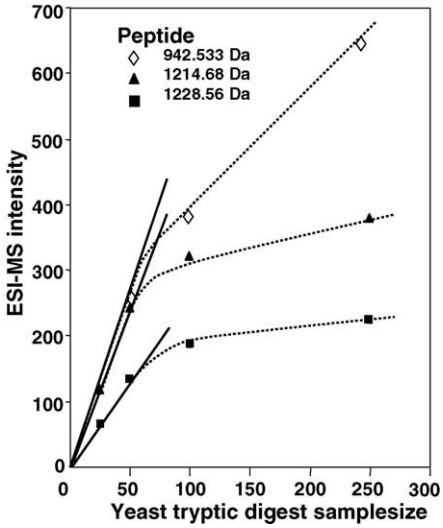


Fig. 15. ESI-MS peak intensities vs. the total proteome sample size for three of the more abundant peptides having the indicated molecular weights. The results were obtained using a 30 μm i.d. packed capillary [42].

50 μm i.d. pre-column (coupled to a 86 cm long, 15 μm i.d. analytical column) can load a 10 μL sample in ∼1.5 min [76]. Another challenge is the amount of time involved with packing such long, small i.d. columns; however, once created, these columns remain highly stable for extend periods of time [77]. A possible solution is to use smaller and more uniform particles, which would improve the ease of packing. Other promising alternatives include using open tubular columns, where the stationary phase is coated on the inside capillary wall [78], or monolithic columns where the stationary phase support is polymerized in situ [79].

The 15 μm i.d. ∼85 cm long column, which provides an optimal flow rate of ∼20 nL/min and requires pressures around 10,000 psi, increases the ESI efficiency by as much as 100-fold more than conventional 150 μm i.d. columns. The 15 μm i.d. column also can maintain peak capacities of ∼1000 [75]. In one instance, the number of peptides detected with LC-FTICR increased approximately three-fold by improving the separation peak capacities from ∼100 to ∼1000 [80]. Similarly, improving LC peak capacities from ∼550 to ∼1000 doubled the number of peptides identified with MS/MS [76].

8. Sensitivity and dynamic range for quantitative measurements

The use of smaller i.d. capillary columns considerably improves sensitivity, while also improving the practical dynamic range of measurements when the absolute sample size is constrained. To examine both the sensitivity and the range of relative protein abundances measurable for complex proteomic samples, we examined a tryptic digest of a mixture containing a 10^6:1 difference in protein abundances for two standards (75 femtomoles cytochrome *c*, and 75 zeptomoles bovine serum albumin) and 5 ng of an $^{14}N/^{15}N$-labeled *D. radiodurans* lysate [81] (see Fig. 16). The results show that proteins having *at least* six orders of magnitude difference in relative abundances can be characterized from a single run of complex proteomic samples. In general, our studies demonstrate that integrated peak intensities reflect relative abundances most precisely for lower abundance species. While the ratios of MS peak intensities significantly deviated from the relative protein content during elution of the most abun-

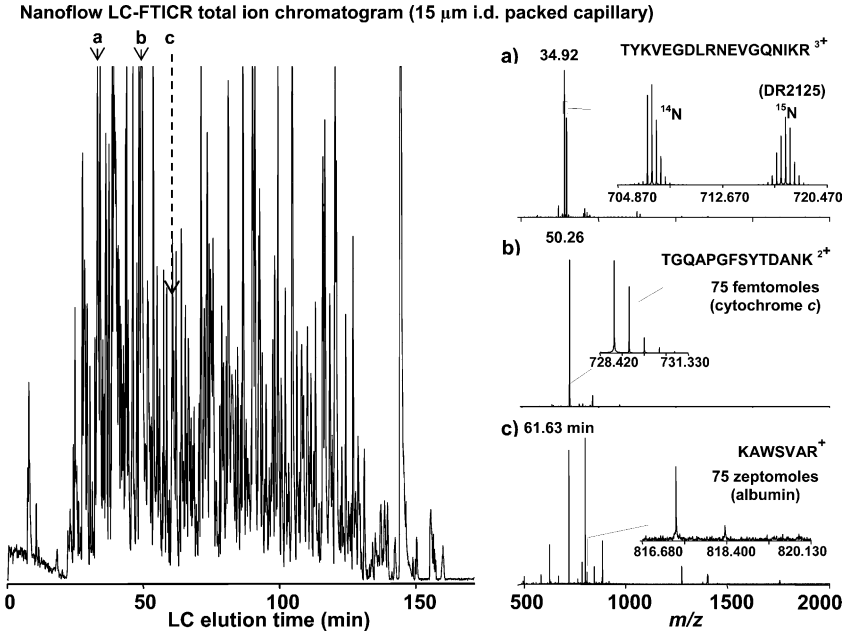


Fig. 16. Examination of the analysis range of *D. radiodurans* relative protein abundances. The sample contained 75 femtomoles cytochrome *c*, 75 zeptomoles bovine serum albumin, and 5 ng of $^{14}N/^{15}N$-labeled *D. radiodurans* tryptic digests.

dant peptides, both this and previous work [77,82] illustrate that quantitative analyses are achievable for other intermediate and lower abundance species. A crucial related point is that high efficiency and high resolution separations allow high abundance species to be generally well separated from most low abundance species, thus enabling better quantitation.

The performance obtainable using a 15 μm i.d. column LC separation in proteome analyses was initially evaluated using micro solid phase extraction (microSPE) and nanoESI coupled on-line to FTICR-MS. Fig. 17 shows the separation of 0.25 ng of a *D. radiodurans* lysate tryptic digest sample using on-line microSPE-nanoLC/nanoESI-FTICR, where a 50 μm i.d. microSPE capillary was used as the pre-loading column [76]. A 10 μl solution of the sample was introduced to the microSPE column at a loading speed of ∼ 8 μl/min at 10,000 psi. Fig. 17 demonstrates the ultra-sensitivity provided by this system that led to identification of 53 *D. radiodurans* ORFs, using $^{15}N/^{14}N$-labeled peak pairs from only 5 pg of *total* sample.

For several years we recognized the potential of an approach that would expand the dynamic range of our MS instrumentation, but implementing it required significant technological developments, as well as fast data-dependent computer control of the experiment. After overcoming several limitations, we developed dynamic range enhancement applied to MS (DREAMS) [83], which is based on the ejection of the most abundant ions in a mass spectrometer so as to provide more efficient use of the FTICR mass spectrometer's dynamic range for each spectrum. To accomplish this task, we developed software that uses the peak intensities from an FTICR mass spectrum to calculate a set of frequencies that are then used to perform dipolar irradiation of ions in an "external" 2D quadrupole to remove the high intensity ions before the collection (or accumulation) step in the external quadrupole. The ions collected in this external quadrupole are then transferred to the FTICR ion trap. Due to elimination of major ions in this way, which would normally result in the rapid filling of the external quadrupole ion accumulation device, longer ion accumulation times can be used to accumulate more of the low abundance ions. The end result of this process is that much greater sensitivity and an extended dynamic range are achieved.

As a demonstration, Fig. 18 displays two partial chromatographic spectra obtained from one LC-FTICR–MS anal-

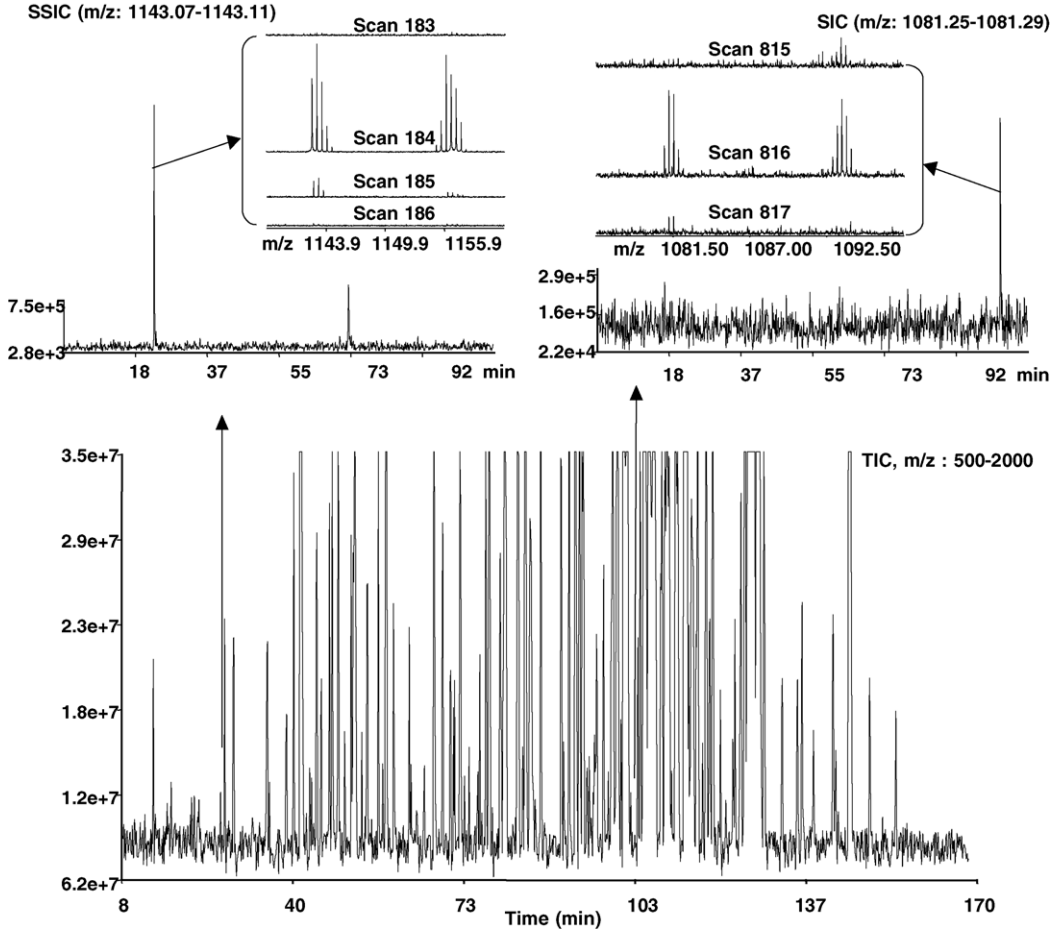


Fig. 17. The ultra-sensitivity provided by microSPE/ nanoLC-FTICR led to identification of 53 *D. radiodurans* ORFs in both of their $^{15}N/^{14}N$-labeled versions from 5 pg of *total* sample. For single peptides, the mass sensitivity can be as high as 10 zeptomole with the concentration detection limit extending to 250 attomolar *concentration* based on injection of the sample in 40 μL solution (which can be loaded in 5 min) and is highly amenable to future automation [56].

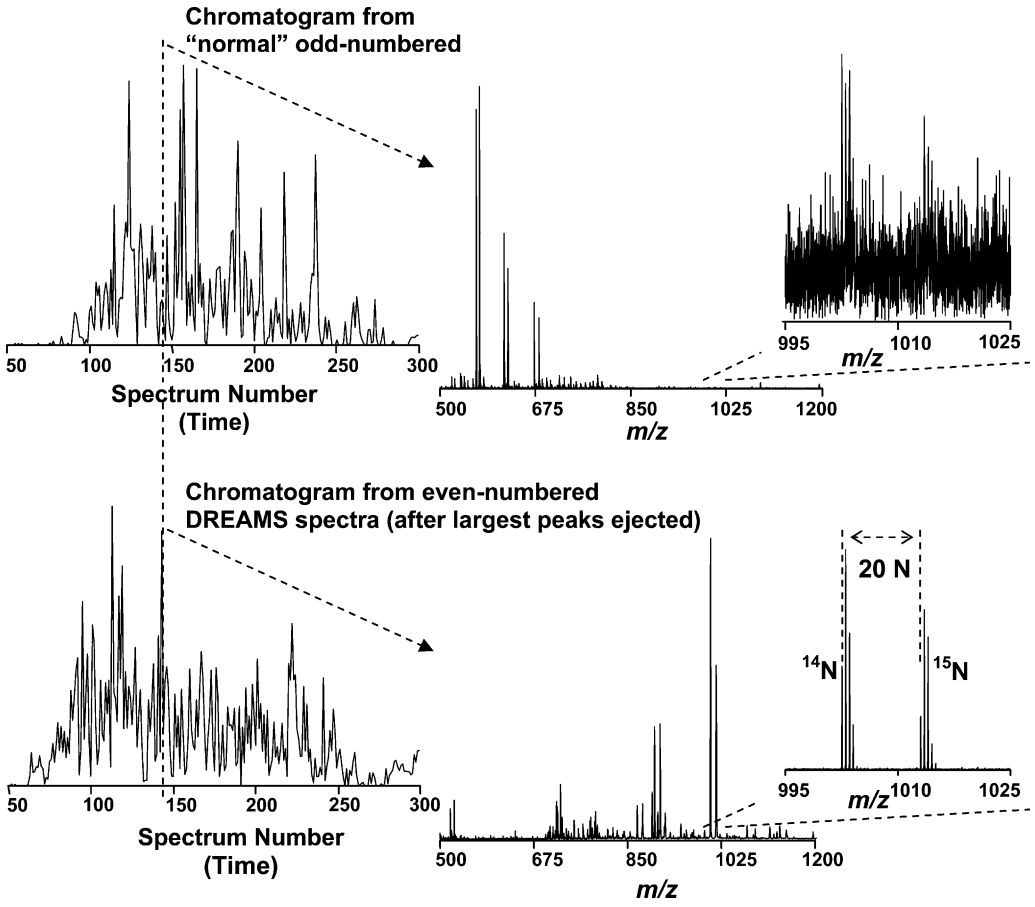


Fig. 18. Partial chromatograms and examples of typical "normal" and DREAMS spectra from a capillary LC-FTICR analysis of peptides from a tryptic digest of a mixture of natural isotopic abundance and ^{15}N-labeled mouse B16 proteins. Top left: partial chromatogram reconstructed from the normal FTICR mass spectra. Bottom left: corresponding chromatogram from the DREAMS spectra for which high relative abundance species were ejected, allowing longer ion accumulation. The mass spectra (center) show the effective ejection of the major species in the top spectrum compared to the one shown on the bottom. The detail (right) shows a large gain in sensitivity and S/N for a peptide pair providing a basis for quantitative comparison of protein abundances in the two cell cultures [62].

ysis, using an equal quantity of ^{15}N and ^{14}N labeled B16 mouse cells [62]. Two different ion chromatograms were reconstructed from this experiment, the first corresponding to the "normal" odd numbered mass spectra and the second to the DREAMS "even" numbered spectra in which the higher abundant ions were ejected. Observed is the removal of the high abundance ions, with a subsequence increase in detection of lower abundant ions. The identification of clear ^{15}N/^{14}N pairs, previously unseen in the "normal" spectra, correlates to a large improvement in sensitivity and dynamic range that can be applied within a single LC-FTICR–MS experiment. By implementing the DREAMS approach, we have typically seen proteomic identification rates increase by 35% [83]. In an application involving a ^{15}N/^{14}N labeled *D. radiodurans* sample, the number of detected peptide pairs increased by ∼50%, and the total number of proteins identified (1244 proteins, representing 40% of the predicted proteome) included 279 proteins detected only in the DREAMS set of spectra [84].

An especially challenging system for detectible dynamic range is the human blood plasma proteome, which is an immensely complex sample with a large dynamic range of relative protein concentrations. More than 99% of the protein content in human plasma is due to only 22 proteins with the most abundant protein, human serum albumin, representing at least half of the total content [85]. To date, we have confidently identified >2000 proteins in human blood plasma samples using our capillary LC separation system coupled with extensive pre-MS separation described above (Shen et al., submitted for publication). The expanded sensitivity and dynamic range afforded by the separation allowed numerous lower abundant species to be detected. Such a detectable dynamic range is needed for the future identification of potential biomarkers for diagnostic purposes.

9. Future directions

The field of proteomics continues to advance, increasingly driven by enhanced MS instrumentation, computational technologies, and quantitative methodologies. As capabilities

continue to mature, they will push the limits of dynamic range detection, efficiency, and quantitation, while providing faster analyses with more reproducibility, and result in a parallel increase in data production. Improvements in the data pipeline, e.g., improved storage, processing, and analysis techniques, will be needed to continue to support increased data generation.

The need for confident, reproducible studies in proteomics for biological applications will continue to drive efforts to improve the reproducibility, and hence quantitation, of repeated analyses for statistical confidence in the results. Standardizing conditions for sample processing and separations will help control and overcome the inherent causes of variances in proteomic measurements. Improving electrospray ionization efficiencies by decreasing the flow rate by the use of smaller capillary i.d.'s is expected to increase both the sensitivity of measurements and provide a basis for improved quantitation. Better LC separations already have been shown to provide large improvements in sensitivity, quantitation and proteome coverage. The use of monolithic columns and microfabricated devices may contribute to future improvements in the realm of nanoLC separations. Other efforts will seek to improve the dynamic range of measurements, which will lead to significantly expanded proteome coverage due to better detection of lesser abundant species.

Such technology advances are leading proteomics towards analysis of smaller samples sizes, with a potential application being proteome analyses of single cells. In other areas of clinical biology, the capability is expected for surveying several hundred small microbiopsy samples and obtaining quantitative information from each to compare populations, characterize disease states, or for time course studies. Proteomic characterization of blood plasma samples may lead to the development of specific disease biomarkers (e.g. for cancer, sepsis) of utility for diagnostic purposes. However, additional advances are needed to address challenges associated with the detection and quantitation of very low level proteins. Based on its performance to date, high efficiency nano-LC coupled with accurate mass instrumentation promises to address the challenges for quantitative high throughput proteomic studies.

Acknowledgments

We thank the U.S. Department of Energy (DOE) Office of Biological and Environmental Research and the NIH National Center for Research Resources (Grant RR018522) for supporting portions of this research. We also thank the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by DOE and located at Pacific Northwest National Laboratory (PNNL) for use of instrumentation. PNNL is operated by Battelle for the DOE under Contract No. DE-AC06-76RLO 1830.

References

[1] M.R. Wilkins, K.L. Williams, R.D. Appel, D.F. Hochstrasser (Eds.), Proteome Research: New Frontiers in Functional Genomics, 1997.
[2] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 9390.
[3] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Anal. Chem. 68 (1996) 850.
[4] M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, M. Mann, Nature 379 (1996) 466.
[5] W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, Proc. Natl. Acad. of Sci. U.S.A. 90 (1993) 5011.
[6] D.J. Pappin, P. Hojrup, A. Bleasby, J. Curr. Biol. 3 (1993) 327.
[7] M. Mann, P. Hojrup, P. Roepstorff, Biol. Mass Spectrom. 22 (1993) 338.
[8] P. James, M. Quadroni, E. Carafoli, G. Gonnet, Biochem. Biophys. Res. Commun. 195 (1993) 58.
[9] J.R. Yates, S. Speicher, P.R. Griffin, T. Hunkapiller, Anal. Biochem. 214 (1993) 397.
[10] K.A. Cox, J.D. Williams, R.G. Cooks, R.E. Kaiser, Biol. Mass Spectrom. 21 (1992) 226.
[11] M.J. Huddleston, M.F. Bean, S.A. Carr, Anal. Chem. 65 (1993) 877.
[12] K. Jonscher, G. Currie, A.L. McCormack, J.R. Yates, Rapid Commun. Mass Spectrom. 7 (1993) 20.
[13] J.A. Loo, C.G. Edmonds, R.D. Smith, Science 248 (1990) 201.
[14] J.A. Loo, C.G. Edmonds, R.D. Smith, Anal. Chem. 63 (1991) 2488.
[15] R.D. Smith, J.A. Loo, C.G. Edmonds, C.J. Barinaga, H.R. Udseth, Anal. Chem. 62 (1990) 882.
[16] A.J. Tomlinson, S. Naylor, J. Liq. Chromatogr. 18 (1995) 3591.
[17] A.L. McCormack, D.M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin, J.R. Yates, Anal. Chem. 69 (1997) 767.
[18] J.R. Yates, A.L. McCormack, J. Eng, Anal. Chem. 68 (1996) A534.
[19] A. Ducret, I. Vanoostveen, J.K. Eng, J.R. Yates, R. Aebersold, Protein Sci. 7 (1998) 706.
[20] A.J. Link, L.G. Hays, E.B. Carmack, J.R. Yates, Electrophoresis 18 (1997) 1314.
[21] J.R. Yates, J. Mass Spectrom. 33 (1998) 1.
[22] M.P. Washburn, D. Wolters, J.R. Yates, Nat. Biotechnol. 19 (2001) 242.
[23] H. Liu, D. Lin, J.R. Yates III, BioTechniques 32 (2002) 898.
[24] D.A. Wolters, M.P. Washburn, J.R. Yates, Anal. Chem. 73 (2001) 5683.
[25] L. Andersson, J. Porath, Anal. Biochem. 154 (1986) 250.
[26] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt, F.M. White, Nat. Biotechnol. 20 (2002) 301.
[27] W. Weckwerth, L. Willmitzer, O. Fiehn, Rapid Commun. Mass Spectrom. 14 (2000) 1677.
[28] M.B. Goshe, T.P. Conrads, E.A. Panisko, N.H. Angell, T.D. Veenstra, R.D. Smith, Anal. Chem. 73 (2001) 2578.
[29] Y. Oda, T. Nagasu, B.T. Chait, Nat. Biotechnol. 19 (2001) 379.
[30] H. Zhou, J.D. Watts, R. Aebersold, R. Nat. Biotechnol. 19 (2001) 375.
[31] W.J. Qian, M.B. Goshe, D.G. Camp, L.R. Yu, K. Tang, R.D. Smith, Anal. Chem. 75 (2003) 5441.
[32] F.Y. Meng, B.J. Cargile, S.M. Patrie, J.R. Johnson, S.M. McLoughlin, N.L. Kelleher, Anal. Chem. 74 (2002) 2923.
[33] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, J. Am. Chem. Soc. 120 (1998) 3265.
[34] H.J. Cooper, R.R. Hudgins, K. Hakansson, A.G. Marshall, J. Am. Soc. Mass Spectrom. 13 (2002) 241.
[35] Y. Ge, B.G. Lawhorn, M. ElNaggar, E. Strauss, J.H. Park, T.P. Begley, F.W. McLafferty, J. Am. Chem. Soc. 124 (2002) 672.
[36] L. Pasa-Tolic, M.S. Lipton, C. Masselon, G.A. Anderson, Y. Shen, N. Tolic, R.D. Smith, J. Mass Spectrom. 37 (2002) 1185.
[37] R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra, H.R. Udseth, Proteomics 2 (2002) 513.

[38] R.D. Smith, L. Pasa-Tolic, M.S. Lipton, P.K. Jensen, G.A. Anderson, Y. Shen, T.P. Conrads, H.R. Udseth, R. Harkewicz, M.E. Belov, C. Masselon, T.D. Veenstra, Electrophoresis 22 (2001) 1652.
[39] R.D. Smith, G.A. Anderson, C. Masselon, M.S. Lipton, L. Pasa-Tolic, T.P. Conrads, T.D. Veenstra, in: B.N. Pramanik, A.K. Ganguly, M.L. Gross (Eds.), Applied Electrospray Mass Spectrometry, 2002, p. 307.
[40] J.P. Speir, M.W. Senko, D.P. Little, J.A. Loo, F.W. McLafferty, J. Mass Spectrom. 30 (1995) 39.
[41] M.V. Gorshkov, L. Pasa-Tolic, H.R. Udseth, G.A. Anderson, B.M. Huang, J.E. Bruce, D.C. Prior, S.A. Hofstadler, L. Tang, L. Chen, J.A. Willett, A.L. Rockwood, M.S. Sherman, R.D. Smith, J. Am. Soc. Mass Spectrom. 9 (1998) 692.
[42] Y. Shen, N. Tolic, C. Masselon, L. Paša-Tolic, I.I.D.G. Camp, M.S. Lipton, G.A. Anderson, R.D. Smith, Anal. Bioanal. Chem. 378 (2004) 1037.
[43] M.E. Belov, R. Zhang, E.F. Strittmatter, D.C. Prior, K. Tang, R.D. Smith, Anal. Chem. 75 (2003) 4195.
[44] J.K. Eng, A.L. McCormack, J.R. Yates, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[45] D.M. Horn, R.A. Zubarev, F.W. McLafferty, J. Amer. Soc. Mass Spectrom. 11 (2000) 320.
[46] K.K. Anderson, M.E. Monroe, D.S. Daly, Proceedings of the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences, 2004, p. 151.
[47] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6591.
[48] T.P. Conrads, K. Alving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Pasa-Tolic, H.R. Udseth, W.B. Chrisler, B.D. Thrall, R.D. Smith, Anal. Chem. 73 (2001) 2132.
[49] L. Pasa-Tolic, P.K. Jensen, G.A. Anderson, M.S. Lipton, K.K. Peden, S. Martinovic, N. Tolic, J.E. Bruce, R.D. Smith, J. Am. Chem. Soc. 121 (1999) 7949.
[50] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[51] S.P. Gygi, B. Rist, R. Aebersold, Biotechnology 11 (2000) 396.
[52] X.D. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Anal. Chem. 73 (2001) 2836.
[53] T. Liu, W.J. Qian, E.F. Strittmatter, D.G. Camp, G.A. Anderson, B.D. Thrall, R.D. Smith, Anal. Chem. 76 (2004) 5345.
[54] I. Stewart, T. Thomson, D. Figeys, Rapid Commun. Mass Spectrom. 15 (2001) 2456.
[55] X. Yao, C. Afonso, C.J. Fenselau, Proteome Res. 2 (2003) 147.
[56] M. Heller, H. Mattou, C. Menzel, X. Yao, J. Am. Soc. Mass Spectrom. 14 (2003) 704.
[57] Y. Shen, J.M. Jacobs, I.I.D.G. Camp, R. Fang, R.J. Moore, R.D. Smith, W. Xiao, R.W. Davis, R.G. Tompkins, Anal Chem. 76 (2004) 1134.
[58] W.J. Qian, J.M. Jacobs, D.G. Camp II, M.E. Monroe, R.J. Moore, M.A. Gritsenko, S.E. Calvano, S.F. Lowry, W. Xiao, L.L. Moldawer, R.W. Davis, R.G. Tompkins, R.D. Smith, Proteomics, in press.
[59] D. Chelius, P.V. Bondarenko, J. Proteome Res. 1 (2002) 317.
[60] P.V. Bondarenko, D. Chelius, T.A. Shaler, Anal. Chem. 74 (2002) 4741.
[61] M.E. Belov, G.A. Anderson, M.A. Wingerd, H.R. Udseth, K. Tang, D.C. Prior, K.R. Swanson, M.A. Buschbach, E.F. Strittmatter, R.J. Moore, R.D. Smith, J. Am. Soc. Mass. Spectrom. 15 (2004) 212.
[62] R.D. Smith, Y. Shen, K. Tang, Acc. Chem. Res. 37 (2004) 269.
[63] L. Tang, P. Kebarle, Anal. Chem. 65 (1993) 3654.
[64] N.B. Cech, C.G. Enke, Mass Spectrom. Rev. 20 (2001) 362.
[65] K. Tang, A. Gomez, Phys. Fluids 6 (1994) 2317.
[66] A. Bruins, Mass Spectrom. Rev. 10 (1991) 53.
[67] J.A. Olivares, N.T. Nguyen, C.R. Yonker, R.D. Smith, Anal. Chem. 59 (1987) 1230.
[68] J.H. Wahl, D.R. Goodlett, H.R. Udseth, R.D. Smith, Anal. Chem. 64 (1992) 3194.
[69] R.D. Smith, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, Anal. Chem. 65 (1993) A574.
[70] D.C. Gale, R.D. Smith, Rapid Commun. Mass Spectrom. 7 (1993) 1017.
[71] M.S. Wilm, M. Mann, Int. J. Mass Spectrom. Ion Process. 136 (1994) 167.
[72] J. Fernandez de la Mora, I. Loscertales, J. Fluid Mech. (1994) 155.
[73] M. Wilm, M. Mann, Anal. Chem. 68 (1996) 1.
[74] A. Schmidt, M. Karas, T.J. Dülcks, Am. Soc. Mass Spectrom. 23 (2003) 492.
[75] Y. Shen, R. Zhao, S.J. Berger, G.A. Anderson, N. Rodriguez, R.D. Smith, Anal. Chem. 74 (2002) 4235.
[76] Y. Shen, R.J. Moore, R. Zhao, J. Blonder, D.L. Auberry, C. Masselon, L. Pasa-Tolic, K.K. Hixson, K.J. Auberry, R.D. Smith, Anal. Chem. 75 (2003) 3596.
[77] Y. Shen, N. Tolic, C. Masselon, L. Pasa-Tolic, D.G. Camp, K.K. Hixson, R. Zhao, G.A. Anderson, R.D. Smith, Anal. Chem. 76 (2004) 144.
[78] P.P.H. Tock, G. Stegeman, R. Peerboom, H. Poppe, J.C. Kraak, K.K. Unger, Chromatographia 24 (1987) 617.
[79] A.R. Ivanov, L. Zang, B.L. Karger, Anal. Chem. 75 (2003) 5306.
[80] Y. Shen, R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, R.D. Smith, Anal. Chem. 73 (2001) 1766.
[81] Y. Shen, N. Tolic, C. Masselon, L. Pasa-Tolic', I.I.D.G. Camp, M.S. Lipton, G.A. Anderson, R.D. Smith, Anal. Bioanal. Chem. 378 (2004) 1037.
[82] R. Zhang, F.E.J. Regnier, Proteome Res. 1 (2002) 139.
[83] M.E. Belov, G.A. Anderson, N.H. Angell, Y. Shen, N. Tolic, H.R. Udseth, R.D. Smith, Anal. Chem. 73 (2001) 5052.
[84] L. Pasa-Tolic, R. Harkewicz, G.A. Anderson, N. Tolic, Y. Shen, R. Zhao, B. Thrall, C. Masselon, R.D. Smith, J. Am. Soc. Mass Spectrom. 13 (2002) 954.
[85] N.L. Anderson, N.G. Anderson, Mol. Cell. Proteomics 1 (2002) 845.
[86] W.J. Qian, D.G. Camp, R.D. Smith, Exp. Rev. Proteomics 1 (1) (2004) 89.